

VIPERdb: a relational database for structural virology

Craig M. Shepherd, Ian A. Borelli, Gabriel Lander, Padmaja Natarajan,
Vinay Siddavanahalli¹, Chandrajit Bajaj¹, John E. Johnson, Charles L. Brooks III
and Vijay S. Reddy*

Department of Molecular Biology, TPC-6 The Scripps Research Institute, 10550 North Torrey Pines Road,
La Jolla, CA 92037, USA and ¹Computational Visualization Center, University of Texas at Austin,
Austin, TX 78712, USA

Received August 31, 2005; Accepted September 23, 2005

ABSTRACT

VIPERdb (<http://viperdbscripps.edu>) is a database for icosahedral virus capsid structures. Our aim is to provide a comprehensive resource specific to the needs of the structural virology community, with an emphasis on the description and comparison of derived data from structural and energetic analyses of capsids. A relational database implementation based on a schema for macromolecular structure makes the data highly accessible to the user, allowing detailed queries at the atomic level. Together with curation practices that maintain data uniformity, this will facilitate structural bioinformatics studies of virus capsids. User friendly search, visualization and educational tools on the website allow both structural and derived data to be examined easily and extensively. Links to relevant literature, sequence and taxonomy databases are provided for each entry.

INTRODUCTION

Even the simplest of viruses represents an elegant and enormous macromolecular assembly, minimally consisting of a nucleic acid genome surrounded by a highly symmetrical protein coat called a capsid. These capsids are described by their T-number, conveniently thought of as the number of protein chains occupying an icosahedral asymmetric unit. For T-numbers greater than 1, the coat protein subunits engage in slightly different interactions depending on their location in the capsid and local symmetry, a phenomenon first described by Caspar and Klug called quasi-equivalence (1).

Continued improvements in instrumentation and computer technologies have resulted in increasing numbers of high- to medium-resolution capsid structures or models obtained from

X-ray crystallography and cryoelectron microscopy analyses, respectively. This growing amount of structural information greatly facilitates the understanding of complex processes in the viral life cycle and could promote the design of new drugs for viral diseases. In addition to its obvious biological and medical importance, these structural data will also prove useful in the general study of self-assembling molecular systems and in developing useful nanomaterials through modification of virus capsids as display platforms for molecules of interest (2,3).

The number and quality of icosahedral virus capsid structures currently in the Protein Data Bank (PDB) makes structural bioinformatics analysis of these assemblies feasible and meaningful. In order to make this task easier a website repository of virus capsid structures, the Virus Particle Explorer (VIPER) was established, and provided data on 74 icosahedral virus capsid structures, representing 21 families and 30 different genera (4). One of the main advantages of the resource was and continues to be the implementation of a consistent convention for orienting the capsid structures in Cartesian space with respect to the icosahedral axes of symmetry. Adopting a single orientation and convention results in a single standard set of matrices for the generation of partial and complete capsid structures from the asymmetric unit deposited in the PDB and facilitates analyses requiring comparisons among diverse capsid structures [e.g. quantifying quasi-equivalence (5) and protein–protein interactions (6)].

However, the storage of data as flat files on a web server offered limited possibilities for data access and made data management and tool development labor intensive. Furthermore, only a non-redundant subset of capsid structures in the PDB were made available. VIPERdb is a new implementation of the resource built on a relational database structure. Together with a concomitant expansion to include all known icosahedral virus structures, it provides a significantly improved resource for the user.

*To whom correspondence should be addressed. Tel: +1 858 784 8191; Fax: +1 858 784 8688; Email: reddiv@scripps.edu

DATABASE STRUCTURE

VIPERdb is implemented as a relational database using the MySQL open source database management system. The structure of the database is based on the mmCIF dictionary using the open source OpenMMS Toolkit (<http://openmms.sdsc.edu>), a collection of software for translating biomolecular structure data in the mmCIF format to a variety of formats, including SQL relational database expressions. As a result, VIPERdb contains tables (~180) for almost every category defined in the mmCIF dictionary. Table columns correspond to data items in the dictionary and provide information ranging from crystallization parameters (e.g. space group) to secondary structure content and topology. Additional tables have been added to store general information on the virus capsids (e.g. name, family, genus and T-number), the matrices required to orient the capsid in the VIPER convention and the derived data from structural and energetic analyses. A schematic diagram showing the main elements and flow of data/information in VIPERdb is shown in Figure 1. The greatest advantage of VIPERdb over the older VIPER implementation is the storage of structural coordinates and related information in the database: this greatly facilitates mining of the data, either for generating dynamic web content on the website or for more large-scale, non-interactive analyses using programming languages with MySQL APIs (e.g. Perl).

DATA CURATION

There are currently a total of 219 entries in VIPERdb, representing 39 and 26 virus genera and families, respectively. New entries are identified by an in-house script, which identifies any structures recently released from the PDB relating to viruses. If a new icosahedral capsid structure is found, the data from its mmCIF file are added using the OpenMMS database loader PDBase. The transformation matrix required to orient the capsid in VIPER orientation is

automatically determined (program available for download from the website) and stored in the database. Some further superficial modifications (e.g. specifying alternate chain IDs) may be made in order to impose consistency with related structures, and are explained more elaborately in the online Supplementary Data. It should be noted that these modifications leave the original data from the PDB intact in the database, and are explicitly noted in REMARK records when a user downloads 'VIPER coordinates' from the website. After deposition of the coordinate data, structural and energetic analyses on the capsid structure are performed using the CHARMM molecular modeling program (7). The results of these analyses are then loaded into specific tables in the database.

DATA ACCESS

Open access to the VIPERdb (<http://viperdb.scripps.edu>) is explained in the main menu on every page at the website (Help, Developer Info). Also included are links to resources explaining how to write applications in a number of languages offering an MySQL API. A simple Perl script which converts PDB coordinates into VIPER coordinates is offered as an example and is available for download along with documentation. For those users unfamiliar with database applications, we have created the 'Database Browser' tool (under 'Utilities' in the main menu). The user can easily examine the tables in the database: mousing over a table name gives a description of the data stored there in a language comprehensible to anyone familiar with the macromolecular structure. Clicking on a table name shows the columns it contains, along with data types. In a similar manner, mousing over the column name describes the data stored there. A query window is provided on every page. The user can choose to have the results returned in an HTML table, or formats compatible with their favorite spreadsheet program. With this tool, the power of the database is readily apparent: answering a question like 'What alanine

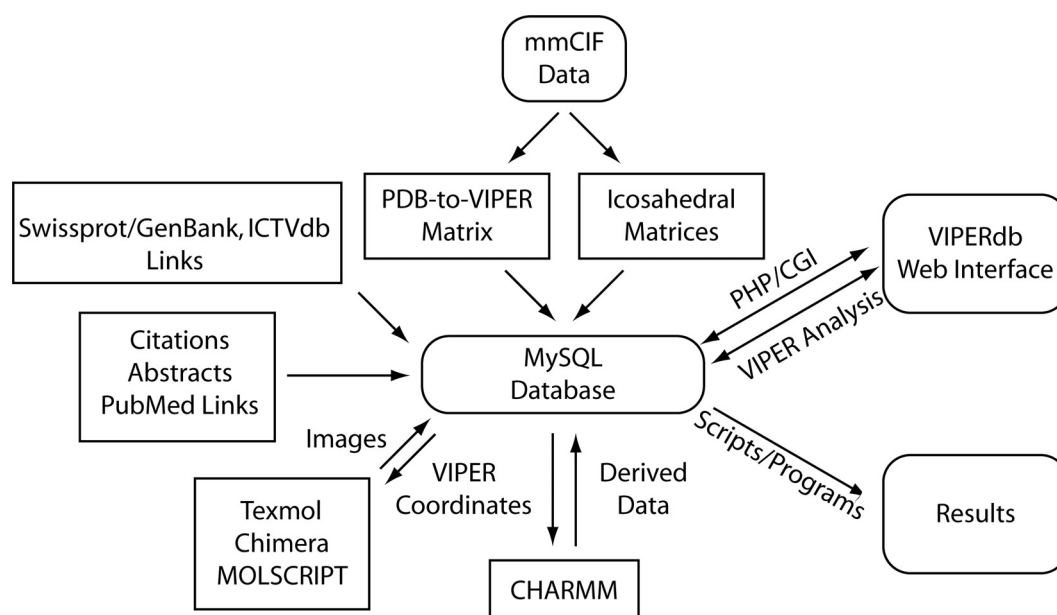


Figure 1. A schematic illustration of flow of data and information in the VIPERdb.

residues make contact with a tryptophan residue in capsids belonging to the family *Picornaviridae*, and at what interfaces?" requires one query.

Anticipating the fact that the majority of users will access VIPERdb through the web interface, significant aesthetic and functional improvements have been made to the website. Although the web pages are now dynamically generated by accessing the database through PHP and Perl scripts, users familiar with VIPER will find that all the data from the older site is still available, and in a more accessible and dynamic format. The easiest way to obtain information on a specific virus capsid is through its 'information page', an example of which is shown in Figure 2. In the case of complex viruses such as double-layered viruses (e.g. reoviruses), separate analyses and web pages were created for individual shells.

The main table at the top of the page consists of three columns, the first of which contains general information about the virus capsid such as PDB ID, structure resolution, links to entries in sequence databases for the capsid coat protein, T-number, number of subunits, surface net charge, and average and maximum diameter. The family and genus of the virus is given, with hyperlinks to the relevant taxon at the ICTVdb (8). Also included are links to separate pages, which provide crystallographic information (space group, crystallization conditions, etc.), primary citation in PubMed and other relevant references, complete with abstracts. The second column of the table includes links to useful information such as the original PDB header information, the PDB-to-VIPER matrix and a link to the transformed VIPER coordinates. Also provided are links to the powerful 'Oligomer Generator' and 'Map a Residue' tools. The former can be used to generate full or partial capsid structures from the asymmetric unit, whereas the latter provides VRML objects highlighting a specific residue of interest in the capsid coat protein asymmetric unit. Pull-down menus in this column also list entries related by genus or family. The last column of the table consists of links to the derived structural data calculated by the VIPERdb team: residue contact tables, association energies, surface accessibility profiles and Q-scores (5). Each set of data is associated with visualization tools at the website including interactive graphs with zoom-in windows and HTML tables with helpful mouseover popups and hyperlinks. These features allow the user to make cross-references between the different types of derived data. Guided examples using these tools and many others are provided in the online Supplementary Data.

The remainder of the information page is devoted to various high-quality renderings of the virus capsid. These include a depth-colored surface volume rendering using TexMol (9), a program which exploits hardware acceleration of widely available graphics cards to rapidly generate 3D structures of large molecular complexes. Users can view these surfaces interactively by downloading a plugin which streams pre-prepared data to their browser. Another pair of renderings, both created with the program MOLSCRIPT (10), aid the viewer in understanding the spatial organization of coat protein chains in the capsid. The first is a representation of the entire capsid, with individual coat protein chains displayed as tubes and color coded according to chain ID. These are shown along with the icosahedral lattice, which divides the capsids up into its

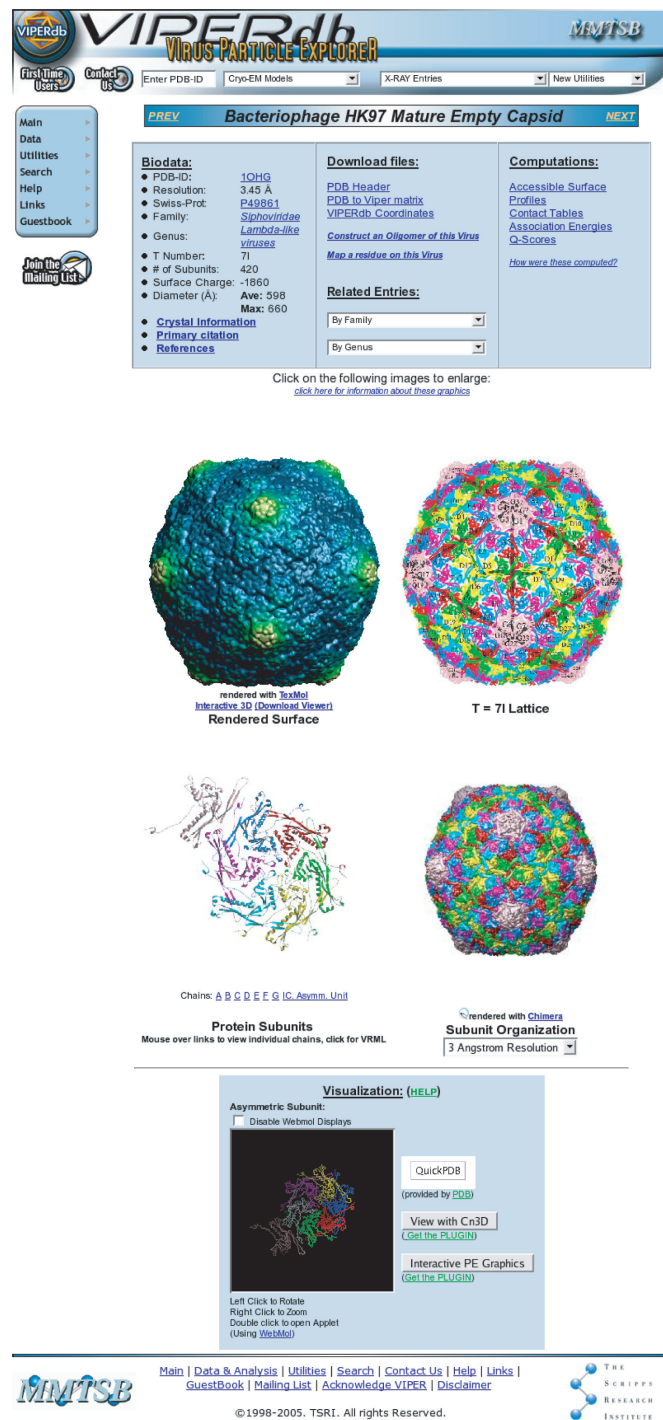


Figure 2. Information page for Bacteriophage HK97 mature empty capsid (PDB ID 10HG). See text for details.

individual asymmetric units. Each chain is labeled with both its chain ID, and a number indicating which of the 60 standard VIPER icosahedral matrices transform the PDB asymmetric unit to this location in the capsid. The other MOLSCRIPT rendering is a close-up view of the capsid asymmetric unit, again with individual chains color coded. Each chain can be individually viewed by mousing over a link below the image. Clicking on the image displays a VRML object which can be

rotated and translated interactively. Another set of highly informative graphics have been made with the Chimera program (7). Different views are selected from a pulldown menu and include surface representations at various grid spacings as well as a combined view, which displays the one asymmetric unit in ribbon mode within a semi-transparent surface. A particularly useful image is the 'Inside view', which provides a perspective from the center of the virus capsid toward the inner surface.

Finally, for the user interested in a quick interactive view of the asymmetric unit structure, a WebMol (11) window is provided at the bottom of the page. Users can select a checkbox to enable/disable the WebMol display, or click on links to view the structure in other freely available viewers, such as Quick-PDB, Cn3D (12) and Protein Explorer (13).

DATABASE UTILITIES AND SERVICES

A number of new tools (see below) were afforded by the new database implementation and the value of the derived data from the structural and energetic analyses.

- (i) Database Browser and SQL query window: learn the structure of VIPERdb and test your own queries.
- (ii) Developer Tools: learn how to access VIPERdb outside of a browser and download a sample Perl script.
- (iii) Contact Finder: locate inter-residue contacts in virus capsids via a search interface.
- (iv) Gallery Maker: assemble a collection of images of chosen capsid entries.
- (v) PDB-to-VIPER: automatically determine the PDB-to-VIPER matrix for your capsid structure.
- (vi) Amino Acid Info: quickly find all data for a specific residue in an entry.

In addition all the VIPER analyses on user-submitted capsid structures (coordinates) are provided as a service.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

We gratefully acknowledge the help and guidance of Dr Tom Goddard from the University of California at San Francisco on using the molecular graphics program Chimera. We also

acknowledge the help of Rido Park of University of Texas at Austin in setting up the TexMol Plugins. VIPERdb is a training/service and dissemination component of the NIH Research Resource: Multiscale Modeling Tools for Structural Biology (MMTSB), which is fully funded by the National Center for Research Resources of the National Institutes of Health (RR12255). Funding to pay the Open Access publication charges for this article was provided by NCRR, RR12255.

Conflict of interest statement. None declared.

REFERENCES

1. Caspar,D.L.D. and Klug,A. (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.*, **27**, 1.
2. Chatterji,A., Burns,L.L., Taylor,S.S., Lomonosoff,G.P., Johnson,J.E., Lin,T. and Porta,C. (2002) Cowpea mosaic virus: from the presentation of antigenic peptides to the display of active biomaterials. *Intervirology*, **45**, 362–370.
3. Chatterji,A., Ochoa,W.F., Paine,M., Ratna,B.R., Johnson,J.E. and Lin,T. (2004) New addresses on an addressable virus nanoblock; uniquely reactive Lys residues on cowpea mosaic virus. *Chem. Biol.*, **11**, 855–863.
4. Reddy,V.S., Natarajan,P., Okerberg,B., Li,K., Damodaran,K.V., Morton,R.T., Brooks,C.L. and Johnson,J.E. (2001) Virus Particle Explorer (VIPER), a website for virus capsid structures and their computational analyses. *J. Virol.*, **75**, 11943–11947.
5. Damodaran,K.V., Reddy,V.S., Johnson,J.E. and Brooks,C.L. (2002) A general method to quantify quasi-equivalence in icosahedral viruses. *J. Mol. Biol.*, **324**, 723–737.
6. Shepherd,C.M. and Reddy,V.S. (2005) Extent of protein–protein interactions and quasi-equivalence in viral capsids. *Proteins*, **58**, 472–477.
7. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
8. Buchen-Osmond,C. (2003) The universal virus database ICTVdb. *Comput. Sci. Eng.*, **5**, 16–25.
9. Bajaj,C., Djeu,P., Siddavanahalli,V. and Thane,A. (2004) Interactive Visual Exploration of Large Flexible Multi-component Molecular Complexes. In *Proceedings of the Annual IEEE Visualization Conference*, October 2004, Austin, Texas, IEEE Computer Society Press, pp. 243–250.
10. Kraulis,P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
11. Walther,D. (1997) WebMol—a Java-based PDB viewer. *Trends Biochem. Sci.*, **22**, 274–275.
12. Hogue,C.W. (1997) Cn3D: a new generation of three-dimensional molecular structure viewer. *Trends Biochem. Sci.*, **22**, 314–316.
13. Martz,E. (2002) Protein Explorer: easy yet powerful macromolecular visualization. *Trends Biochem. Sci.*, **27**, 107–109.